

Novelty and Collective Attention

Fang Wu and Bernardo A. Huberman

Information Dynamics Laboratory

HP Labs

Palo Alto, CA 94304

February 5, 2008

Abstract

The subject of collective attention is central to an information age where millions of people are inundated with daily messages. It is thus of interest to understand how attention to novel items propagates and eventually fades among large populations. We have analyzed the dynamics of collective attention among one million users of an interactive website — digg.com — devoted to thousands of novel news stories. The observations can be described by a dynamical model characterized by a single novelty factor. Our measurements indicate that novelty within groups decays with a stretched-exponential law, suggesting the existence of a natural time scale over which attention fades.

The problem of collective attention is at the heart of decision making and the spread of ideas, and as such it has been studied at the individual and small group level by a number of psychologists [1, 2], economists [3], and researchers in the area of marketing and advertising [4, 5, 6]. Attention also affects the propagation of information in social networks, determining the effectiveness of advertising and viral marketing [7]. And while progress on this problem has been made in small laboratory studies and in the theoretical literature of attention economics [8], it is still lacking empirical results from very large groups in a natural, non-laboratory, setting.

To understand the process underlying attention in large groups, consider as an example how a news story spreads among a group of people. When it first comes out, the story catches the attention of a few ones, who may further pass it on to others if they find it interesting enough. If a lot of people start to pay attention to this story, its exposure in the media will continue to increase. In other words, a positive reinforcement effect sets in such that the more popular the story becomes, the faster it spreads.

This growth is counterbalanced by the fact that the novelty of a story tends to fade with time and thus the attention that people pay to it. Therefore, in considering the dynamics of collective attention two competing effects are present: the growth in the number of people that attend to a given story and the habituation that makes the same story less likely to be attractive as time goes on. This process becomes more complex in the realistic case of multiple items or stories appearing at the same time, for now people also have the choice of which stories to attend with their limited attention.

In order to study the dynamics of collective attention and its relation to novel inputs in a natural setting, we analyzed the behavioral patterns of one million people interacting with a news website whose content is solely determined by its own users. Because people using this website assign each news story an explicit measure of popularity, we were able to determine the growth and decay of attention for thousands of news stories and to validate a theoretical model which predicts both the dynamics and the statistical distribution of story lifetimes.

The website under study, **digg.com**, is a digital media democracy which allows its users to submit news stories they discover from the Internet [9]. A new submission immediately appears on a repository webpage called “Upcoming Stories”, where other members can find the story and, if they like it, add a *digg* to it. A so-called *digg number* is shown next to each story’s headline, which simply counts how many users have *digg*ed the story in the

past.¹ If a submission fails to receive enough diggs within a certain time limit, it eventually falls out of the “Upcoming” section, but if it does earn a critical mass of diggs quickly enough, it becomes *popular* and jumps to the **digg.com** front page.² Because the front page can only display a limited number of stories, old stories eventually get replaced by newer stories as the page gets constantly updated. If a story however, becomes very popular it qualifies as a “Top 10” and stays on the right side of the front page for a very long time.

When a story first appears on the front page it attracts much attention, causing its digg number, N_t , to build up quickly. After a couple of hours its digg rate slows down because of both its lack of novelty and its lack of prominent visibility (reflected in the fact that it moves away from the front page). Thus the digg number of each story eventually saturates to a value, N_∞ , that depends on both its popularity growth and its novelty decay. In order to determine the statistical distribution of this saturation number, which corresponds to the number of diggs it has accumulated throughout its evolution, we measured the histogram of the final diggs of all 29,864 popular stories in the year 2006. As can be seen from Fig. 1, the distribution appears to be quite skewed, with the normal Q-Q plot of $\log(N_\infty)$ a straight line. A Kolmogorov-Smirnov normality test of $\log(N_\infty)$ with mean 6.546 and standard deviation 0.6626 yields a p -value of 0.0939, suggesting that N_∞ follows a log-normal distribution.

It is then natural to ask whether N_t , the number of diggs of a popular story after finite time t , also follows a log-normal distribution. To answer this question, we tracked the digg numbers of 1,110 stories in January 2006 minute by minute. The distribution of $\log(N_t)$ again obeys a bell shape curve. As an example, a Kolmogorov-Smirnov normality test of $\log(N_{2 \text{ hours}})$ with mean 5.925 and standard deviation 0.5451 yields a p -value as high as 0.5605, supporting the hypothesis that N_t also follows a log-normal distribution.

The log-normal distribution can be explained by a simple stochastic dynamical model which we now describe. If N_t represents the number of people who know the story at time t , in the absence of any habituation,

¹In fact, digg users are given the option to “bury” a story, which will *decrease* the story’s digg number. Because this rarely happens, we ignore this possibility and simply assume that a story’s digg number can only grow with time.

²The actual machine-learning algorithm used to determine whether a story qualifies to appear on the front page is very complex and will not be discussed in this paper [10].

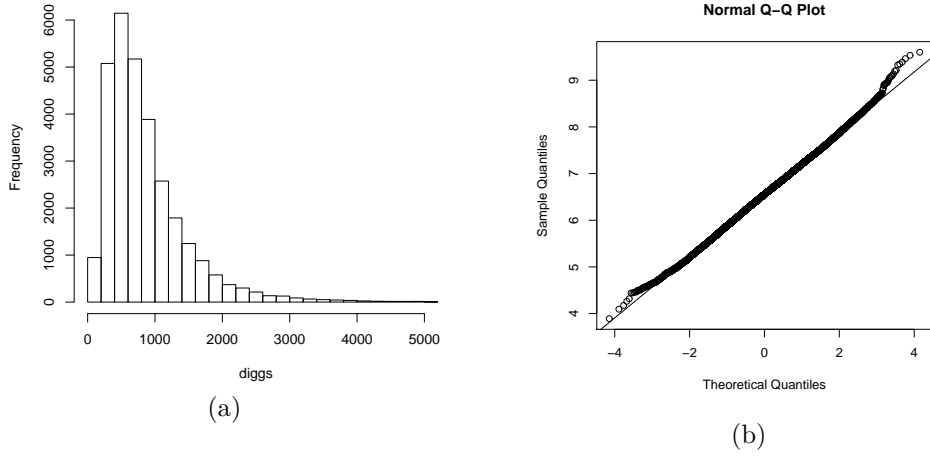


Figure 1: (a) The histogram of the 29,684 diggs in 2006, as on January 9, 2007. (b) The normal Q-Q plot of $\log(N_\infty)$. The straight line shows that $\log(N_\infty)$ follows a normal distribution with a slightly longer tail. This is due to **digg.com**’s built-in reinforcement mechanism that favors those “top stories”, which can stay on the front page and can be found at many other places (e.g. “popular stories in 30 days” and “popular stories in 365 days”).

on average a fraction μ of those people will further spread the story to some of their friends. Mathematically this assumption can be expressed as $N_t = (1 + X_t)N_{t-1}$, where X_1, X_2, \dots are positive i.i.d. random variables with mean μ and variance σ^2 . The requirement that X_i must be positive ensures that N_t can only grow with time. As we have discussed above, this growth in time is eventually curtailed by a decay in novelty, which we parameterize by a time dependent *factor* r_t consisting of a series of decreasing positive numbers with the property that $r_1 = 1$ and $r_t \downarrow 0$ as $t \uparrow \infty$. With this additional parameter, the full stochastic dynamics of story propagation is governed by $N_t = (1 + r_t X_t)N_{t-1}$, where the factor $r_t X_t$ acts as a discounted random multiplicative factor. When X_t is small (which is the case for small time steps) we have the following approximate solution:

$$N_t = \prod_{s=1}^t (1 + r_s X_s) N_0 \approx \prod_{s=1}^t e^{r_s X_s} N_0 = e^{\sum_{s=1}^t r_s X_s} N_0, \quad (1)$$

where N_0 is the initial population that is aware of the story. Taking logarithm of both sides, we obtain

$$\log N_t - \log N_0 = \sum_{s=1}^t r_s X_s. \quad (2)$$

The right hand side is a discounted sum of random variables, which for r_t near one (small time steps) can be shown to be described by a normal distribution [11]. It then follows that for large t the probability distribution of N_t will be approximately log-normal.

Our dynamic model can be further tested by taking the mean and variance of both sides of Eq. (2):

$$\frac{E(\log N_t - \log N_0)}{\text{var}(\log N_t - \log N_0)} = \frac{\sum_{s=1}^t r_s \mu}{\sum_{s=1}^t r_s \sigma^2} = \frac{\mu}{\sigma^2}. \quad (3)$$

Hence if our model is correct, a plot of the sample mean of $\log(N_t) - \log(N_0)$ versus the sample variance for each time t , should yield a straight line passing through the origin with slope μ/σ^2 . One such plot for 1,110 stories collected in January 2007 is shown in Fig. 2. As can be seen, the points indeed lie on a line with slope 6.947.

The decay factor r_t can now be computed explicitly from N_t up to a constant scale. Since we have normalized r_1 to 1, we have

$$r_t = \frac{E(\log N_t) - E(\log N_{t-1})}{E(\log N_1) - E(\log N_0)}. \quad (4)$$

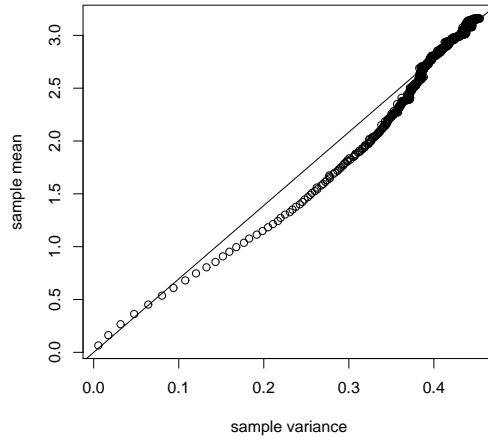


Figure 2: Sample mean of $\log N_t - \log N_0$ versus sample variance, for 1,110 stories in January 2007. Time unit is one minute. The points are plotted as follows. For each story we calculate the quantity $\log N_t - \log N_0$, which is the logarithm of its digg number measured t minutes after its first appearance on the front page, minus the logarithm of its initial digg number. We collect 1,110 such quantities for 1,110 stories. Then we compute their sample mean y and sample variance x , and mark the point (x, y) . This is for one t . We repeat the process for $t = 1, 2, \dots, 1440$ and plot 1440 points in total (i.e. 24 hours). They lie roughly on a straight line passing through the origin with slope 6.947.

The curve of r_t estimated from the 1,110 stories in January 2007 is shown in Fig. 3(a). As can be seen, r_t decays very fast in the first two to three hours, and its value becomes less than 0.03 after three hours. Fig. 3(b,c) show that r_t decays slower than exponential and faster than power law. Fig. 3(d) shows that r_t can be fit empirically to a *stretched exponential relaxation* or Kohlrausch-Williams-Watts law [12]: $r_t \sim e^{-0.4t^{0.4}}$. The *halflife* τ of r_t can then be determined by solving the equation

$$\int_0^\tau e^{-0.4t^{0.4}} dt = \frac{1}{2} \int_0^\infty e^{-0.4t^{0.4}} dt. \quad (5)$$

A numerical calculation gives $\tau = 69$ minutes, or about one hour. This characteristic time is consistent with the fact that a story usually lives on the front page for a period between one and two hours.

The stretched exponential relaxation often occurs as the result of multiple characteristic relaxation time scales [12, 13]. This is consistent with the fact that the decay rate of a story on **diggs.com** depends on many factors, such as the story's topic category, the time of a day when it appears on the front page. The measured decay factor r_t is thus an average of these various rates and describes the collective decay of attention.

These measurements, comprising the dynamics of one million users attending to thousands of novel stories, allowed us to determine the effect of novelty on the collective attention of very large groups of individuals, while nicely isolating both the speed of propagation of new stories and their decay. We also showed that the growth and decay of collective attention can be described by a dynamical model characterized by a single novelty factor which determines the natural time scale over which attention fades. The exact value of the decay constant depends on the nature of the medium but its functional form is universal. These experiments, which complement large social network studies of viral marketing [7] are facilitated by the availability of websites that attract millions of users, a fact that turns the internet into an interesting natural laboratory for testing and discovering the behavioral patterns of large populations on a very large scale [14].

References

- [1] Kahneman, D. Attention and effort. Englewood Cliffs, N.J.: Prentice Hall (1973).

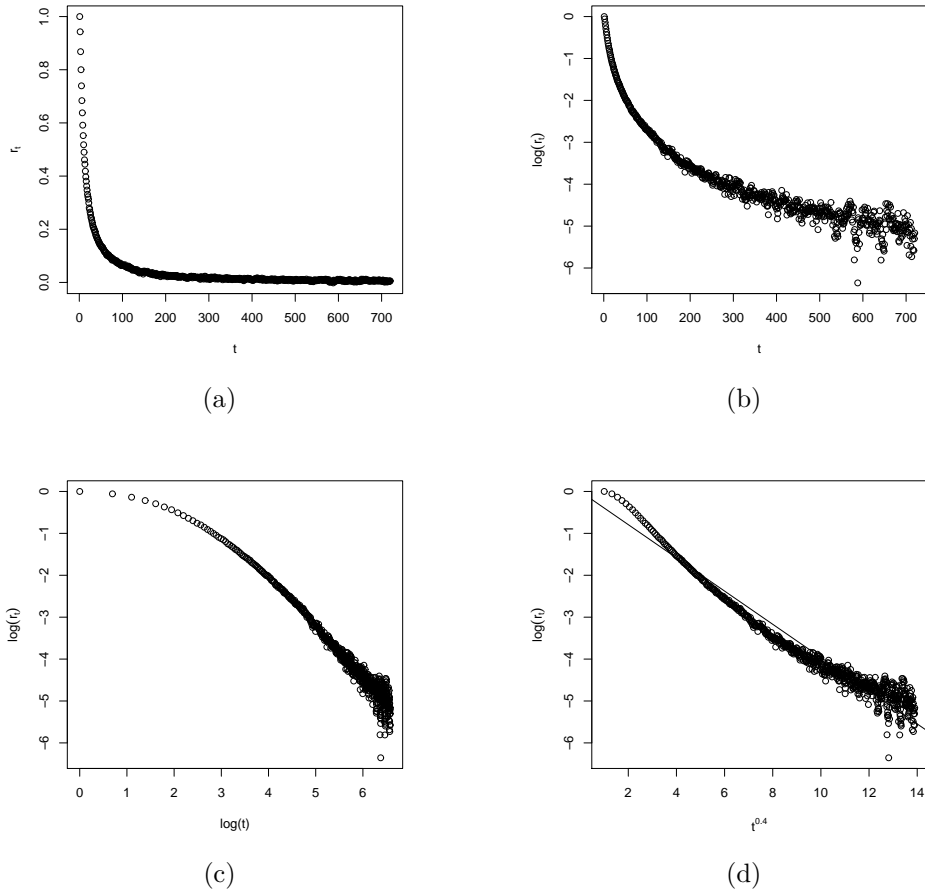


Figure 3: (a) The decay factor r_t as a function of time. Time t is measured in minutes. (b) $\log(r_t)$ versus t . r_t decays slower than exponential. (c) $\log(r_t)$ versus t . r_t decays faster than power law. (d) $\log(r_t)$ versus $t^{0.4}$. The slope is approximately -0.4 .

- [2] Pashler, H. E. The psychology of attention. MIT Press (1998).
- [3] Camerer, C. The behavioral challenge to economics: understanding normal people. Paper presented at Federal Reserve of Boston meeting (2003).
- [4] Pieters, F. G. M., Rosbergen, E. and Wedel, M. Visual attention to repeated print advertising: A test of scanpath theory. *Journal of Marketing Research* 36(4), 424–438 (1999).
- [5] Dukas, R. Causes and consequences of limited attention. *Brain Behavior and Evolution* 63, 197–210 (2004).
- [6] Reis, R. Inattentive Consumers. *Journal of Monetary Economics* Vol. 53, 1761–1800 (2006).
- [7] Lefkovic, J., Adamic, L. and Huberman, B. A. The dynamics of viral marketing. *Proceedings of the ACM Conference on Electronic Commerce* (2006).
- [8] Falkinger, J. Attention Economies. Forthcoming in *Journal of Economic Theory* (2003).
- [9] How Digg Works. <http://www.digg.com/how>
- [10] Private communication with the digg.com support team.
- [11] Embrechts, P. and Maejima, M. The central limit theorem for summability methods of i.i.d. random variables. *Probability Theory and Related Fields*, Volume 68, Number 2 (1984).
- [12] Lindsey, C. P. and Patterson, G. D. Detailed comparison of the William-Watts and Cole-Davidson functions. *J. Chem. Phys.* **73**(7) (1980).
- [13] Frisch, U. and Sornette, D. Extreme deviations and applications. *J. Phys. I France* **7** 1155–1171 (1997).
- [14] Watts, D. A twenty first century science. *Nature* **445**, pp. 489 (2007).